

10-2011

## Noncanonical Hydrogen Bonding In Nucleic Acids. Benchmark Evaluation Of Key Base-phosphate Interactions In Folded Rna Molecules Using Quantum-chemical Calculations And Molecular Dynamics Simulations

Marie Zgarbova

Petr Jurecka

Pavel Banas

Michal Otyepka

Judit E. Sponer

*See next page for additional authors*

Follow this and additional works at: [https://scholarworks.bgsu.edu/chem\\_pub](https://scholarworks.bgsu.edu/chem_pub)

 Part of the [Chemistry Commons](#)

---

### Repository Citation

Zgarbova, Marie; Jurecka, Petr; Banas, Pavel; Otyepka, Michal; Sponer, Judit E.; Leontis, Neocles B.; Zirbel, Craig L.; and Sponer, Jiri, "Noncanonical Hydrogen Bonding In Nucleic Acids. Benchmark Evaluation Of Key Base-phosphate Interactions In Folded Rna Molecules Using Quantum-chemical Calculations And Molecular Dynamics Simulations" (2011). *Chemistry Faculty Publications*. 51.  
[https://scholarworks.bgsu.edu/chem\\_pub/51](https://scholarworks.bgsu.edu/chem_pub/51)

This Article is brought to you for free and open access by the Chemistry at ScholarWorks@BGSU. It has been accepted for inclusion in Chemistry Faculty Publications by an authorized administrator of ScholarWorks@BGSU.

---

**Author(s)**

Marie Zgarbova, Petr Jurecka, Pavel Banas, Michal Otyepka, Judit E. Sponer, Neocles B. Leontis, Craig L. Zirbel, and Jiri Sponer

# Noncanonical Hydrogen Bonding in Nucleic Acids. Benchmark Evaluation of Key Base–Phosphate Interactions in Folded RNA Molecules Using Quantum-Chemical Calculations and Molecular Dynamics Simulations

Marie Zgarbová,<sup>†</sup> Petr Jurečka,<sup>†</sup> Pavel Banáš,<sup>†</sup> Michal Otyepka,<sup>†</sup> Judit E. Šponer,<sup>†,§</sup> Neocles B. Leontis,<sup>||</sup> Craig L. Zirbel,<sup>||</sup> and Jiří Šponer<sup>\*,†,§</sup>

<sup>†</sup>Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University, Olomouc, tr. 17. listopadu 12, 771 46 Olomouc, Czech Republic

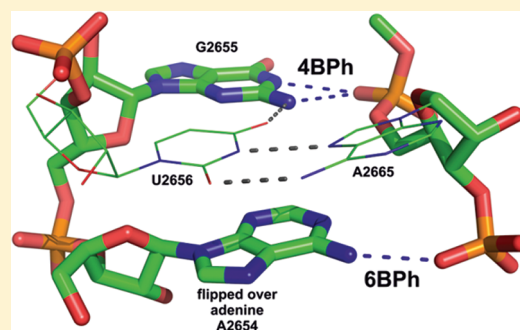
<sup>‡</sup>Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 61265 Brno, Czech Republic

<sup>§</sup>CEITEC - Central European Institute of Technology, Masaryk University, Žerotínovo nám. 9, 60177 Brno, Czech Republic

<sup>||</sup>Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States

## S Supporting Information

**ABSTRACT:** RNA molecules are stabilized by a wide range of noncanonical interactions that are not present in DNA. Among them, the recently classified base–phosphate (BPh) interactions belong to the most important ones. Twelve percent of nucleotides in the ribosomal crystal structures are involved in BPh interactions. BPh interactions are highly conserved and provide major constraints on RNA sequence evolution. Here we provide assessment of the energetics of BPh interactions using MP2 computations extrapolated to the complete basis set of atomic orbitals and corrected for higher-order electron correlation effects. The reference computations are compared with DFT-D and DFT-D3 approaches, the SAPT method, and the molecular mechanics force field. The computations, besides providing the basic benchmark for the BPh interactions, allow some refinements of the original classification, including identification of some potential doubly bonded BPh patterns. The reference computations are followed by analysis of some larger RNA fragments that consider the context of the BPh interactions. The computations demonstrate the complexity of interaction patterns utilizing the BPh interactions in real RNA structures. The BPh interactions are often involved in intricate interaction networks. We studied BPh interactions of protonated adenine that can contribute to catalysis of hairpin ribozyme, the key BPh interaction in the S-turn motif of the sarcin–ricin loop, which may predetermine the S-turn topology and complex BPh patterns from the *glmS* riboswitch. Finally, the structural stability of BPh interactions in explicit solvent molecular dynamics simulations is assessed. The simulations well preserve key BPh interactions and allow dissection of structurally/ functionally important water-mediated BPh bridges, which could not be considered in earlier bioinformatics classification of BPh interactions.



## INTRODUCTION

Folded RNA molecules are stabilized by a variety of intermolecular interactions involving nucleobases as well as ribose and phosphates from the RNA backbone. The structural and functional variability of RNA molecules primarily originates from noncanonical interactions, i.e., all interactions other than those occurring in canonical double helices. The exact shape (stericity) of the noncanonical interactions and motifs is of primary importance for RNA function and evolution.<sup>1</sup> Although many of the noncanonical interactions have already been systematically characterized with experimental, bioinformatic, and theoretical approaches,<sup>1–23</sup> the base–phosphate (BPh) hydrogen bonds have only recently been classified.<sup>24</sup>

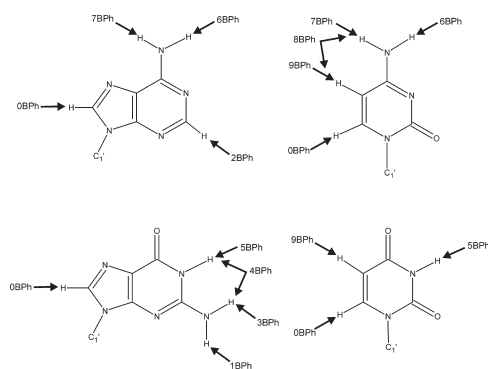
BPh interactions are ubiquitous in large functional RNAs and ribonucleoprotein particles. Approximately 12% of the nucleotides in the ribosome form direct internucleotide H-bonds between nucleobase donor atoms and phosphate oxygen acceptor atoms.<sup>24,25</sup> BPh interactions contribute to stabilization of key recurrent RNA structural motifs such as the UUCG and GNRA hairpin tetraloops<sup>26–28</sup> and the sarcin–ricin internal loop.<sup>24,29,30</sup>

**Special Issue:** Pavel Hobza Festschrift

**Received:** May 24, 2011

**Revised:** August 15, 2011

**Published:** September 12, 2011



**Figure 1.** Basic classification of BPh interactions. The arrows identify the atoms that are involved in direct base-phosphate H-bonds.<sup>24</sup>

BPh interactions are sequence-specific and represent a major constraint on the evolution of RNA sequences.<sup>24</sup> Thus, BPh interactions belong to the core of the biochemically and evolutionarily most important molecular interactions shaping RNA structures.

The classification of the BPh interactions in RNA has been derived by combining structural bioinformatics and quantum-chemical (QM) calculations.<sup>24</sup> The classification considers the distribution of phosphate groups around nucleobases in experimental structures without including the broader structural context of the interactions and possible water-mediated substates. On the basis of the distribution of phosphate groups around the nucleobases, 17 different BPh binding patterns have been suggested for the four nucleobases, which could be assigned to 10 different binding types labeled as 0BPh–9BPh (Figure 1). The BPh interactions utilize the H-bond donor sites of the nucleobase and one of the anionic oxygens of the phosphate moiety. Occasionally, the esteric oxygen may also participate in the binding in combination with the anionic oxygen. These combined binding modes, such as 4BPh and 8BPh exhibit the highest stability and, accordingly, are highly conserved in RNA architectures.<sup>24</sup> Although the base–base and base–sugar interactions are rarely stabilized by C–H···O H-bonds, such H-bonds are quite frequent among BPh binding patterns: 3 out of the 10 basic BPh types utilize C–H···O H-bonding. This is not surprising, because the phosphate oxygens are better proton acceptors than the formally neutral N and O acceptors of nucleobases.

In the previous study we used simple quantum chemical models to unravel the most fundamental physicochemical differences between the various BPh binding types.<sup>24</sup> Our model included a nucleobase and a phosphate group, both surrounded with a continuum solvent assuming a dielectric constant  $\epsilon = 80$  (at 298 K). The role of the continuum solvent was 2-fold: (i) to screen the electrostatic field of the phosphate group, which would dominate the gas phase calculations to an extent that is not relevant to the solvated RNA molecules, and (ii) to separate weaker BPh binding patterns as genuine local minima on the potential energy surface. Although the calculations were quite simple, the distinct computational minima found gave us a guideline for how to distinguish between the physicochemically different interaction modes of nucleobases and phosphates. In addition, our theoretical interaction energy estimates exhibited a sound correlation with the frequency of occurrences of the various BPh binding patterns in the course of evolution.<sup>24</sup> This means that the stronger the BPh interaction type is, the more likely it is realized during evolution. This illustrates that besides

the shape compatibility, the energetics of interactions also plays an important role in selecting the intermolecular interactions contributing to the stabilization of functional RNA molecules. In many instances, the BPh interactions represent a direct extension of the generalized RNA base pairing classification (the Leontis and Westhof families)<sup>1,5,7</sup> that considers base–base, base–sugar, and sugar–sugar H-bonds. The BPh interactions are the dominant H-bonds in some recurrent base pairing patterns and thus may convert some weak pairing patterns into strong ones.<sup>24</sup> BPh interactions are present also in DNA four-way junctions.<sup>31</sup> Unrelated intranucleotide base-phosphate interactions occurring in model studies of isolated nucleotides were studied using QM calculations.<sup>32,33</sup>

Taking into consideration the importance of the BPh interactions for RNA structure, dynamics, function, and evolution, we extend the preceding computations. We first perform a set of calculations with a more complete model than in the earlier study. Besides optimized structures we also use geometries taken from ribosomal X-ray structures. For selected BPh interactions, where we noticed some salient structural context, we also perform calculations using extended model systems, to better understand the context dependence of the BPh interactions. We also assess behavior of BPh interactions in classical explicit solvent simulations. Finally, we carry out reference (highest-accuracy) computations and test performance of several quantum-chemical methods and molecular mechanics force field at estimating the strength of these key RNA molecular interactions. As the reference method, we use MP2<sup>34</sup> calculations expanded to the complete basis set<sup>35,36</sup> of atomic orbitals supplemented by corrections carried out at the CCSD(T)<sup>37–39</sup> level of theory.

## METHODS

**Initial Geometries.** We have considered two sets of systems. Small model structures for the individual BPh interactions were selected from the crystal structures of ribosomal subunits 1JSE and 1S72.<sup>3,40</sup> Another set of larger models of BPh geometries (from sarcin–ricin loop, hairpin ribozyme, and *glmS* riboswitch) was taken from the MD simulations of the respective RNA systems.<sup>41,42</sup> All initial structures were capped with hydrogen atoms and optimized as described in the Geometry Optimization section. Dimethyl phosphate was used as the small model. The larger models will be specified below. For the larger models we also performed calculations by splitting the interacting monomers into fragments and evaluating the interfragment contributions separately. Such calculations provide better insight into the balance of forces participating in the interaction patterns. The fragments were always prepared as chemically complete monomers, typically capping them by hydrogen atoms.

**Geometry Optimizations.** Geometry optimizations were carried out at the density functional theory (DFT) level with B3LYP<sup>43</sup> density functional with the 6-31G\*\*<sup>44,45</sup> basis set augmented with the sp-type diffuse orbitals for the two anionic oxygens of the phosphate group (one s and one p function, exponent 0.0845) and for the phosphorus (exponent 0.0348). The following constraints were applied during optimizations to keep the relative orientation of the monomers close to the initial X-ray geometries. C3', C5', C1', and P atoms were kept at their initial Cartesian positions (see Supporting Information Figure S1 for atom numbers). These constraints help to fix the experimental geometry while still allowing sufficient flexibility to relax intramolecular degrees of freedom as well as all H-bonds participating

in the BPh interactions. When needed, additional constraints were applied to dihedral angles determining relative orientation of the monomers and occasionally on the intermolecular angles (detailed description of the constraints is given in Supporting Information, Table S1 and Figure S1). Intermolecular distances were thus allowed to relax within the given constraints. For the structures containing protonated adenine A38H<sup>+</sup> from the catalytic pocket of the hairpin ribozyme (HrRz) we used the IEFPCM solvation model<sup>46,47</sup> during optimization (structures HrRz str4, HrRz str5, HrRz str6, and HrRz str7; see below). The geometry optimizations were carried out in the Gaussian03 program<sup>48</sup> except for the three stacked structures from *glmS* riboswitch (*glmS* str1, *glmS* str2, and *glmS* str3) where we used Turbomole 5.10<sup>49,50</sup> program in connection with a code enabling for inclusion of the dispersion term.<sup>51</sup> For the later stacked systems we used the density functional theory (DFT) augmented with empirical dispersion term (DFT-D).<sup>51</sup> We used again B3LYP functional with 6-31G\*\* basis, and the dispersion parametrization was based on radii scaling (exponent 23 and  $s_R = 1.1$ ).<sup>51</sup>

**Interaction Energies.** Interaction energies were calculated as a difference between total energy of the dimer and total energies of monomers in the dimer geometry. Trimers were evaluated analogously.<sup>52</sup> The three-body terms were calculated only for two trimers containing two bases and one phosphate group (i.e., *glmS* str6 and *glmS* str7). In both structures we obtained repulsive vacuum three-body term of ~2.5 kcal/mol. For the other trimers containing a bridging water molecule we assumed that the three-body terms would be negligible.

The interaction energy  $\Delta E_{\text{COSMO}}^{\text{METHOD}}$  has been calculated as sum of in vacuo interaction energy  $\Delta E_{\text{VAC}}^{\text{METHOD}}$  and solvent contribution to interaction (Gibbs energy)  $\Delta \Delta G_{\text{COSMO}}$  using eq 1:

$$\Delta E_{\text{COSMO}}^{\text{METHOD}} = \Delta E_{\text{VAC}}^{\text{METHOD}} + \Delta \Delta G_{\text{COSMO}} \quad (1)$$

where the superscript METHOD denotes either MP2 or CBS(T) method used for in vacuo interaction energy calculation ( $\Delta E_{\text{VAC}}^{\text{METHOD}} = E_{\text{vac}}^{\text{METHOD}}(\text{AB}) - E_{\text{vac}}^{\text{METHOD}}(\text{A}) - E_{\text{vac}}^{\text{METHOD}}(\text{B})$ ; see below).  $\Delta \Delta G_{\text{COSMO}}$  was calculated as a difference between the solvation Gibbs energies of dimer and monomers,  $\Delta \Delta G_{\text{COSMO}} = \Delta G_{\text{COSMO}}(\text{AB}) - \Delta G_{\text{COSMO}}(\text{A}) - \Delta G_{\text{COSMO}}(\text{B})$ , using RIMP2 level of theory and PTE approach.<sup>53</sup>

**i. Nonextrapolated Interaction Energies.** In the case of base–phosphate interactions taken from hairpin ribozyme, sarcin–ricin loop motif, and *glmS* riboswitch, we calculated the in vacuo interaction energies on MP2 level. The  $\Delta E_{\text{VAC}}^{\text{MP2/aDZ}}$  MP2<sup>34</sup> calculations were performed with the RI approximation (RI-MP2)<sup>50</sup> with the aug-cc-pVDZ<sup>54,55</sup> basis set and were corrected for the basis set superposition error (BSSE) using the counterpoise procedure.<sup>56</sup> The solvent Gibbs energy contribution to the interaction,  $\Delta \Delta G_{\text{COSMO}}$ , was calculated using the COSMO<sup>57</sup> continuum model at the RIMP2/aug-cc-pVDZ level of theory as a difference between solvation energies of the dimer and the isolated monomers. Note that in eq 1 the in vacuo interaction energy is combined with the change of the solvation Gibbs energy; therefore, the final quantity on the left-hand side, despite being denoted as  $\Delta E$ , contains some entropic terms coming from solvation. However, we still denote it as interaction energy rather than free energy (although such evaluations are sometimes denoted as  $\Delta G$  in the literature) as it is not equivalent to experimentally measurable free energies. The numbers are best interpreted as interaction energy with inclusion of solvent screening of electrostatic stabilization.

**ii. Extrapolated Interaction Energies Corrected for the Higher-Order Correlation Contributions.** In the case of the partially optimized structures of the small models taken from X-ray structures (1JSE and 1S72), we substantially increased the level of single-point interaction energy computations. The  $\Delta E_{\text{VAC}}^{\text{METHOD}}$  in eq 1 is extrapolated to the complete basis set (CBS) limit at MP2 level and corrected for the higher-order correlation effects by adding a  $\Delta \text{CCSD(T)}$  correction. The RIMP2 calculations were performed with aug-cc-pVDZ (aDZ) and aug-cc-pVTZ (aTZ)<sup>54,55</sup> basis sets and extrapolated according to Helgaker and co-workers (eq 2).<sup>35,36</sup>

$$E^{\text{HF}/X} = E^{\text{HF/CBS}} + A e^{-\alpha X} \quad E^{\text{MP2}_{\text{corr}}/X} = E^{\text{MP2}_{\text{corr/CBS}}} + B X^{-3} \quad (2)$$

$E^{\text{HF}/X}$  and  $E^{\text{MP2}_{\text{corr}}/X}$  are HF energies and MP2 corrections for the basis set with the highest angular momentum  $X$  ( $X = 2$  for aDZ and  $X = 3$  for aTZ), respectively, and  $E^{\text{HF/CBS}}$  and  $E^{\text{MP2}_{\text{corr/CBS}}}$  stand for HF energy and MP2 corrections at the complete basis set limit, respectively. The  $\Delta \text{CCSD(T)}$  correction is the difference between the  $\text{CCSD(T)}^{37-39}$  and MP2 interaction energies calculated in a small 6-31+G\*\*<sup>44,45</sup> basis set. The total in vacuo energy evaluation is given by eq 3. The CBS(T) abbreviation indicates that the  $\text{CCSD(T)}$  calculation has not been extrapolated to CBS. (Such calculations are rather often incorrectly marked as  $\text{CCSD(T)/CBS}$  in the literature.)

$$\begin{aligned} \Delta E_{\text{VAC}}^{\text{CBS(T)}} &= \Delta E_{\text{VAC}}^{\text{HF/CBS}} + \Delta E_{\text{VAC}}^{\text{MP2}_{\text{corr/CBS}}} \\ &+ (\Delta E_{\text{VAC}}^{\text{CCSD(T)_{corr}/6-31+G*}} - \Delta E_{\text{VAC}}^{\text{MP2}_{\text{corr}/6-31+G*}}) \end{aligned} \quad (3)$$

Also the solvation contribution to the interaction energy ( $\Delta \Delta G_{\text{COSMO}}^{\text{solv}}$ ) was evaluated in aDZ and aTZ basis sets and extrapolated to CBS limit (HF and MP2 contributions separately). MP2 calculations were carried out in TurboMole 5.10<sup>49,50</sup> program with RI approximation. The  $\text{CCSD(T)}$  calculations were performed in the Molpro package.<sup>58</sup>

**Decomposition of Interaction Energies Using SAPT.** The interaction energies of the partially optimized structures taken from 1JSE and 1S72 crystal structures were decomposed into components using the DFT-SAPT (symmetry adapted perturbation theory)<sup>59-63</sup> method implemented in the Molpro program package.<sup>58</sup> In DFT-SAPT the monomer is described by density functional theory (DFT) and the intermolecular interactions by SAPT. The total interaction energy is sum of the following terms (eq 4)

$$E^{\text{SAPT}} = E_{\text{el}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} + \delta(\text{HF}) \quad (4)$$

$E^{\text{SAPT}}$  corresponds to the  $\Delta E_{\text{VAC}}^{\text{METHOD}}$  term defined above. In the equation the individual terms mean electrostatic ( $E_{\text{el}}^{(1)}$ ), exchange repulsion ( $E_{\text{exch}}^{(1)}$ ), induction including charge transfer ( $E_{\text{ind}}^{(2)}$ ), and dispersion ( $E_{\text{disp}}^{(2)}$ ) contributions and their mixing terms ( $E_{\text{exch-ind}}^{(2)}$  and  $E_{\text{exch-disp}}^{(2)}$ ). The mixing terms can be added up with their corresponding second-order components, leading to  $E_{\text{ind}}$  and  $E_{\text{disp}}$  terms.  $\delta(\text{HF})$  is a correction term for the higher order contributions. All DFT-SAPT calculations were done with the aug-cc-pVDZ basis set and the asymptotically corrected exchange–correlation density functional PBE0AC<sup>60</sup> combined from the PBE0<sup>64</sup> modification of the PBE xc-functional<sup>65</sup> and the



asymptotically correct LB94xc-potential<sup>66</sup> through the gradient-controlled shift procedure.<sup>67</sup> The gradient-controlled shift procedure requires a difference (shift) between the vertical ionization potential (IP) and HOMO energy of the DFT method as an input. Herein the IPs were calculated at the PBE0/cc-pVTZ level, and the HOMO values were taken from the aug-cc-pVDZ calculation.

**Single Point DFT Calculations with Empirical Dispersion Correction.** DFT-D is a combination of a standard DFT calculation with an empirical correction for the long-range correlation. Here we used for single point calculations a combination of the TPSS<sup>68</sup> density functional, 6-311++G(3df,3pd)<sup>69–72</sup> basis set and dispersion parametrization based on radii scaling<sup>51</sup> (D-0.96-27). We also used a more recent DFT-D3 method by Grimme et al.,<sup>73</sup> with TPSS/TZVPP combination and dispersion parameters optimized for the TZVPP<sup>74</sup> basis set (m4 grid was used). DFT calculations were performed with the TurboMole 5.10 program<sup>49,50</sup> combined with the Fortran code *tm\_disp* for dispersion correction (<http://fch.upol.cz/en/software/>).

**Molecular Mechanics Calculations.** The molecular mechanics (MM) interaction energies were calculated with the sander module of the AMBER<sup>75</sup> program package and the *ff94*<sup>76</sup> set of van der Waals parameters. Point charges were taken from the *ff94* force field and charges on hydrogen atoms of the capping methyl groups were chosen so as to fix integral charge of the molecule (all capping hydrogen atoms equal). Charges on the PO<sub>4</sub><sup>−</sup> groups and carbon atoms of dimethyl phosphate were taken from *ff94*, and all six hydrogen atoms on the methyl groups were assigned a charge of +0.0252 au to fix integral charge of −1 au. Charges for the protonated adenine were taken from ref 42. For some structures we tested a different charge-derivation procedure, as described in the Results. All MM energies were evaluated using the QM optimized geometries.

**Molecular Dynamics (MD) Simulations.** MD simulation trajectories were taken from our preceding studies.<sup>41,42</sup> The simulations were carried out using the AMBER code and versions of the Cornell et al. *ff99* force field.<sup>76</sup> Standard protocols have been used as specified elsewhere.<sup>28,41,42,77,78</sup> The structural stability of BPh interaction was analyzed by measuring of the occupancy of corresponding H-bonds forming the given BPh interaction. The distance cutoff of 3.5 Å (distance between heavy atoms) and angle cutoff of 40° for H-bond angle were used to define the H-bond. Although these cutoffs are widely used for definition of H-bond contact in MD simulations, they are still ad hoc defined and might result in inaccurate estimation of H-bond occupancy due to stochastic fluctuation beside these cutoffs. Thus, we ignored all disruptions of the H-bonds that were shorter than 10 ps by assuming that they reflect genuine stochastic fluctuation of H-bound state. Similarly, the occurrences of H-bond shorter than 100 ps were not taken into account as they most likely did not correspond to the formation of the H-bound state.

## ■ RESULTS AND DISCUSSION

**Basic Characterization of the BPh Interactions.** We first extend the preceding QM computations<sup>24</sup> on BPh interactions to obtain better insight into the nature of BPh interactions and mainly to get a set of reference structure-energy data for benchmarking other computational methods.

A broad set of diverse 32 geometries was selected from the crystal structures of ribosomal subunits 1JSE and 1S72.<sup>3,40</sup> Note

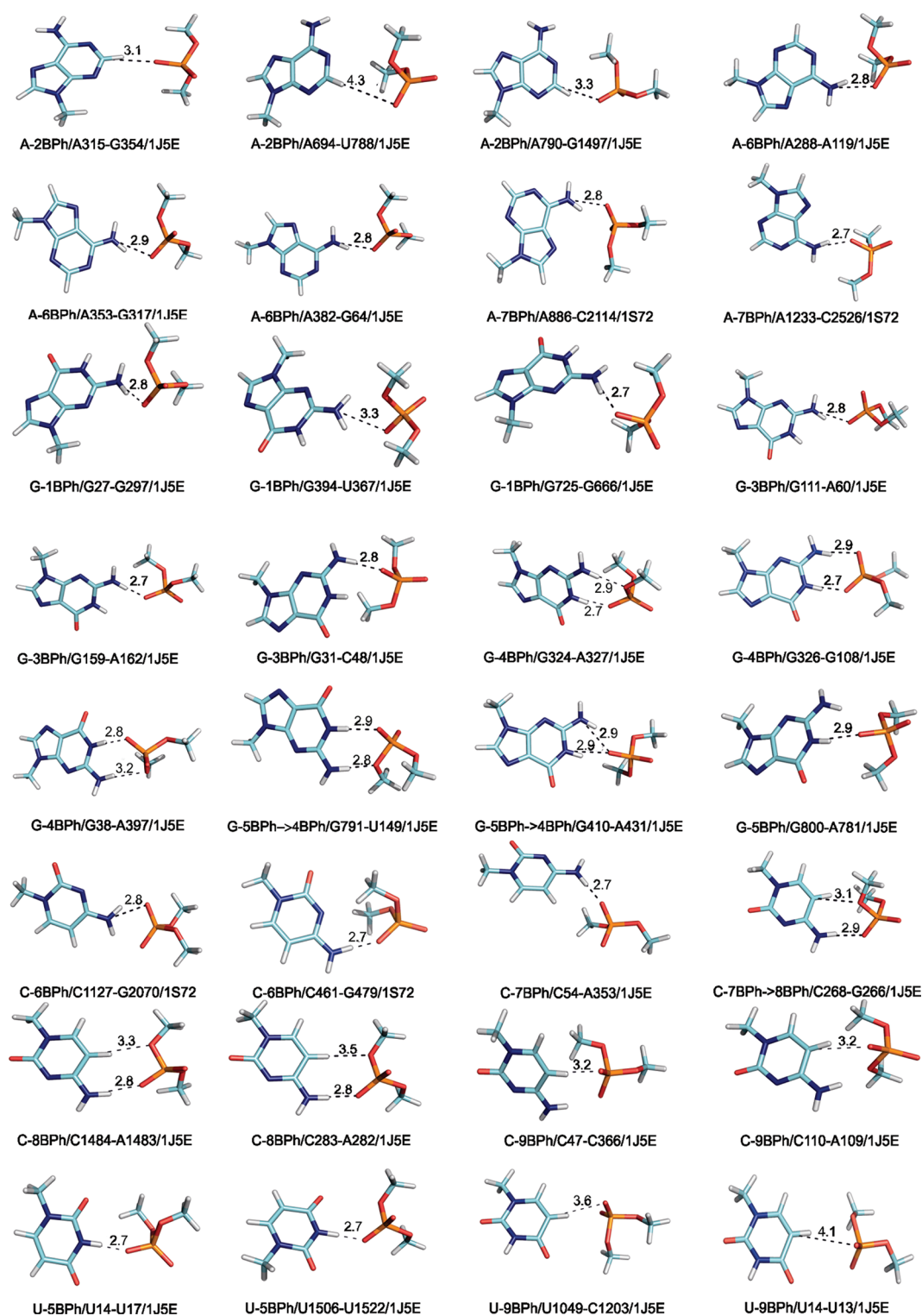
that the angular variability of observed interactions can be partially affected by the limited resolution of the experimental structures. We included all the BPh families except of the 0BPh pattern, which was proposed to be rather opportunistic and less relevant interaction binding motif.<sup>24</sup> We considered 2 or 3 distinct starting structures of each BPh pattern to get some idea about the shape of the potential energy surface of BPh interactions. However, in some cases we had to reclassify the binding pattern mainly due to relaxation and switches of the binding motif during the optimization. Nonetheless, each BPh family is represented by at least one representative structure. In comparison with the previous QM calculations,<sup>24</sup> we extended the model of phosphate to the more realistic dimethyl phosphate and increased the level of QM theory to MP2/CBS corrected to higher-order correlation effects using CCSD(T) correction in small basis set (further denoted as CBS(T) method).

The geometries taken from X-ray were optimized under several constraints to keep the geometry close to the initial X-ray structures (for detailed description of the constraints used, see Methods). The constraints were applied in such a way that the intermolecular distances could partially relax, while the relative monomer orientation was preserved. Still, similar motifs exhibit substantial spread of the intermolecular distances, as for instance in the A-2BPh family (Figure 2).

The abbreviations of the individual structures are constructed in the following way. Their first letter stands for the nucleobase involved in BPh interaction, followed by the BPh family label,<sup>24</sup> X-ray (PDB file) nucleotide numbering of the respective interacting nucleobase and phosphate and, finally, the PDB ID of the structure.

The optimized geometries are very similar to the initial structures, with the exception of two structures that rearranged during optimization from the initial 7BPh and 5BPh geometries to structurally close 8BPh and 4BPh geometries, respectively; in Table 1 they are denoted as C-7BPh→8BPh/C268-G266/1JSE and G-5BPh→4BPh/G410-A431/1JSE. In one case, the initial geometry seemed to belong to a different family than in the original classification<sup>24</sup> and was reclassified (G-5BPh→4BPh/G791-U149/1JSE in Figure 2 and Table 1).

**Bifurcated Interactions.** Several structures do not seem to be unambiguous representatives of a given family and might deserve introduction of a new subfamily. For instance, in the G-1BPh/G394-U367/1JSE structure both hydrogen atoms of the N2 amino group interact with the phosphate oxygen atoms making two hydrogen bonds, one of them typical for the 1BPh family<sup>24</sup> and the other for the 3BPh family. Other structures in which both hydrogen atoms of the amino group are involved in hydrogen bonding interaction are A-6BPh/A353-G317/1JSE, G-1BPh-G27-G297/1JSE, and C-6BPh-C1127-G2070/1S72; thus it concerns all bases having the amino group (A, C, and G). These doubly bonded (bifurcated) structures are energetically favorable in vacuo. Why did they, however, not appear in the original classification?<sup>24</sup> In the analysis of Zirbel et al.<sup>24</sup> the potentially doubly bonded structures appear to be less populated (albeit not absent) than the neighboring structures with a single hydrogen bond to the amino group, despite the fact that the doubly bonded structures are energetically favorable in vacuo. This might be either due to structural constraints imposed by RNA structures or due to specific or nonspecific solvent effects that may alter the local potential energy surface. Due to the obvious resolution limits of the original structural bioinformatics study, it is not possible to confidently assess the individual occurrences of the interactions; i.e., we do not know which of the observed instances reveal real



**Figure 2.** Model structures for the individual base–phosphate interactions (BPh) taken from the crystal structures 1J5E and 1S72. H-bonds are indicated with dashed lines. The structures are designated using the following nomenclature: x-NBPh/xL-yM/ndb, where x stands for the nucleobase donating the H-bond, NBPh stands for the type of BPh interaction as classified by Zirbel et al.,<sup>24</sup> xL-yM lists both participating nucleotides including their numbers based on the respective X-ray structure (the first nucleotide contributes its nucleobase to the interaction, and the second contributes its phosphate), and ndb is the code of the respective X-ray structure.

doubly bonded interactions and which of them represent outliers due to resolution and refinement limits. Nevertheless, it appears that some such interactions do occur in RNAs.

*Energetics of the Interactions.* Figure 2 summarizes the resulting geometries; note that names of the structures reflect the initial classification based on ref 24. As pointed out above, in very few

**Table 1. Total Interaction Energies (kcal/mol) for the Model Structures Taken from 1J5E and 1S72 Crystal Structures Calculated by Different Methods<sup>a</sup>**

structure	MP2/aDZ	MP2/aTZ	MP2/CBS	$\Delta E_{\text{VAC}}^{\text{CBS(T)}}$	$\Delta E_{\text{COSMO}}^{\text{CBS(T)}}$	$\Delta E_{\text{IDEAL}}$	DFT-D	DFT-D3	$E^{\text{SAPT}}$	MM
A-2BPh/A315-G354/1J5E	−3.15	−3.62	−3.80	−4.17	−0.29		−3.48	−3.98	−3.38	5.72
A-2BPh/A694-U788/1J5E	−0.99	−1.19	−1.27	−1.41	0.25	−0.1	−1.41	−1.70	−1.18	0.75
A-2BPh/A790-G1497/1J5E	−2.88	−3.31	−3.49	−3.82	−1.04		−3.23	−3.72	−3.09	1.98
A-6BPh/A288-A119/1J5E	−15.14	−15.81	−16.09	−15.84	−3.77		−16.37	−17.34	−14.50	−10.67
A-6BPh/A353-G317/1J5E	−15.86	−16.43	−16.68	−16.54	−4.45	−3.1	−17.03	−18.00	−15.33	−11.67
A-6BPh/A382-G64/1J5E	−16.22	−16.87	−17.14	−16.97	−4.47		−17.42	−18.39	−15.64	−11.79
A-7BPh/A886-C2114/1S72	−14.33	−14.82	−15.03	−14.13	−1.89	−2.8	−15.26	−16.34	−13.82	−10.52
A-7BPh/A1233-C2526/1S72	−14.89	−15.79	−16.15	−14.25	−3.15		−16.67	−17.61	−14.28	−11.18
G-1BPh/G27-G297/1J5E	−23.66	−24.69	−25.12	−25.06	−5.47		−25.53	−26.71	−23.71	−23.60
G-1BPh/G394-U367/1J5E	−19.45	−20.01	−20.25	−20.38	−2.67	−4.3	−20.78	−21.70	−19.59	−19.72
G-1BPh/G725-G666/1J5E	−22.98	−24.06	−24.51	−24.39	−6.39		−25.66	−26.32	−22.88	−19.67
G-3BPh/G111-A60/1J5E	−26.28	−26.93	−27.19	−27.37	−4.93		−27.48	−28.42	−26.23	−24.81
G-3BPh/G159-A162/1J5E	−27.49	−28.63	−29.08	−29.02	−6.82	−5.4	−29.76	−30.73	−27.20	−24.69
G-3BPh/G31-C48/1J5E	−27.09	−28.34	−28.84	−28.66	−8.14		−29.39	−30.53	−26.59	−25.91
G-4BPh/G324-A327/1J5E	−28.06	−29.47	−30.05	−29.47	−10.32	−8.3	−30.03	−31.15	−26.68	−25.19
G-4BPh/G326-G108/1J5E	−33.98	−35.40	−35.98	−35.78	−10.98		−35.52	−36.78	−33.10	−31.44
G-4BPh/G38-A397/1J5E	−28.74	−29.72	−30.13	−29.80	−9.35		−30.02	−31.19	−27.69	−26.62
G-5BPh→4BPh/G791-U149/1J5E	−26.28	−27.50	−28.00	−27.58	−9.05		−27.79	−29.11	−25.13	−24.33
G-5BPh→4BPh/G410-A431/1J5E	−31.99	−32.93	−33.30	−33.24	−10.27		−32.95	−34.19	−31.01	−30.06
G-5BPh/G800-A781/1J5E	−32.37	−33.28	−33.64	−33.50	−10.89	−4.0	−33.21	−34.57	−31.47	−31.21
C-6BPh/C1127-G2070/1S72	−21.14	−22.15	−22.58	−22.48	−3.90	−3.5	−21.93	−23.14	−20.88	−19.65
C-6BPh/C461-G479/1S72	−15.98	−17.50	−18.10	−17.71	−4.03		−17.77	−18.71	−15.37	−9.63
C-7BPh/C54-A353/1J5E	−25.54	−26.57	−26.97	−26.70	−6.61	−4.8	−26.62	−27.70	−25.12	−22.65
C-7BPh→8BPh/C268-G266/1J5E	−24.06	−24.94	−25.31	−24.89	−6.21	−5.4	−25.01	−26.02	−23.50	−19.94
C-8BPh/C1484-A1483/1J5E	−24.83	−25.90	−26.32	−26.00	−5.65		−25.33	−26.56	−24.40	−22.91
C-8BPh/C283-A282/1J5E	−25.83	−26.82	−27.22	−26.92	−6.83		−26.76	−27.91	−25.40	−23.37
C-9BPh/C47-C366/1J5E	−22.21	−23.41	−23.92	−23.40	−3.35	−0.6	−23.25	−24.32	−22.05	−15.59
C-9BPh/C110-A109/1J5E	−21.05	−21.74	−22.04	−21.62	−3.79		−21.77	−22.62	−20.80	−16.89
U-5BPh/U14-U17/1J5E	−13.63	−14.64	−15.02	−14.75	−6.02	−4.2	−14.76	−15.68	−12.83	−7.79
U-5BPh/U1506-U1522/1J5E	−14.18	−15.26	−15.68	−15.32	−5.07		−15.46	−16.58	−13.41	−6.38
U-9BPh/U1049-C1203/1J5E	−9.04	−9.28	−9.39	−9.38	0.20	−0.6	−9.21	−9.79	−9.13	−6.83
U-9BPh/U14-U13/1J5E	−3.49	−3.60	−3.65	−3.56	1.11		−3.72	−4.07	−3.48	−1.11

<sup>a</sup> All energies are derived for identical QM-optimized geometries. MP2/aDZ, MP2/aTZ = gas phase energies computed by RIMP2/aug-cc-pVDZ and RIMP2/aug-cc-pVTZ methods with correction for the BSSE. MP2/CBS = extrapolated energy to the complete basis set.  $\Delta E_{\text{VAC}}^{\text{CBS(T)}}$  = MP2/CBS energy corrected by adding a  $\Delta \text{CCSD(T)}$  correction.  $\Delta E_{\text{COSMO}}^{\text{CBS(T)}}$  = final interaction energy computed with the COSMO dielectric continuum model. DFT-D = DFT calculation with empirical dispersion correction calculated by TPSS/6-311++G(3df,3pd)/D-0.96-27.<sup>51</sup> DFT-D3 = DFT calculation with D3 empirical dispersion correction calculated by TPSS/TZVPP.<sup>73</sup>  $E^{\text{SAPT}}$  = DFT-SAPT energy calculated by PBE0AC/aug-cc-pVDZ. MM = molecular mechanics interaction energy in vacuum (parm99).  $\Delta E_{\text{IDEAL}}$  = interaction energy derived by Zirbel et al. with inclusion of solvent screening in ref 24.

cases the constraints in fact could not prevent a change of the interaction pattern. Nevertheless, the broad set of 32 structures is more than sufficient for the purpose of reference computations.

Table 1 presents the calculated interaction energies. The first four columns of Table 1 summarize standard gas phase reference computations, i.e., MP2/CBS with subsequent CCSD(T) correction (see eqs 2 and 3 in Methods). We denote this calculation level as CBS(T) rather than CCSD(T)/CBS, to explicitly mark that the higher-order electron correlation contributions are estimated with small basis set. The calculated gas phase interaction energies are in the range −1 to −36 kcal/mol. It confirms that the BPh interactions represent a set of diverse H-bonds spanning from very weak contacts up to very strong hydrogen bond interactions. Deformation energies of the monomers were not included in the present computations, albeit for a representative set of structures they were calculated at the TPSS/LP/D

level and are given in Supporting Information (Table S2). The deformation energies are between 2.5 and 4.4 kcal/mol for those systems evaluated. Such deformation energies are entirely expected, because interactions involving charged phosphate groups are significantly stronger than comparable interactions in neutral molecules. For instance, deformation energy of the canonical GC base pair with three hydrogen bonds is 3.6 kcal/mol at the MP/CBS level with respect to nonplanar (fully optimized) monomers.<sup>79</sup> Further, the DFT method likely overestimates the monomer deformation energies. Indeed, we calculated that the TPSS/cc-pVTZ method gives 4.5 kcal/mol deformation energy of the GC base pair with respect to fully optimized monomers. Thus, deformation energies would not affect ranking of the BPh interactions significantly.

Note that the studied complexes are non-neutral, and thus the gas phase interaction energies should not be trivially compared



with for example neutral base pairs. Interestingly, even the strongest BPh interactions are weaker than the  $\text{CCH}^+$  hemiprotonated base pair ( $-52$  kcal/mol in the gas phase).<sup>79</sup> The  $\text{CCH}^+$  base pair, however, is triply bonded, and in real systems it requires large free energy input to protonate the cytosine, which has an intrinsic  $\text{pK}_a$  of  $\sim 4.2$ .<sup>80</sup> The BPh interactions, in contrast, include NA building blocks in their canonical state.<sup>80</sup> Interaction energies of canonical GC and AT base pairs are  $-32$  and  $-17$  kcal/mol, respectively.<sup>79</sup> Interaction energies of all three above-noted base pairs are calculated by the same procedure as the BPh interactions; i.e., the numbers are given without the addition of deformation energies.

The fifth column of Table 1 shows the interaction energies  $\Delta E_{\text{COSMO}}^{\text{CBS(T)}}$  calculated with inclusion of the continuum solvent correction (see eq 1 in Methods). These calculations represent the basic stability assessment of the interactions inside nucleic acids. For comparison with recently published results the sixth column shows the energies  $\Delta E_{\text{IDEAL}}$  taken from the original study.<sup>24</sup> Note that the original calculations were carried out on idealized (centroid) geometries, using a less complete model system (only the dihydrogen phosphate anion was used, whereas dimethyl phosphate is used here) and were not expanded to CBS. Also, in case of the G-SBPh family one of the hydrogen atoms of the  $\text{NH}_2$  group of guanine was replaced by a methyl group to prevent formation of a more stable G-4BPh binding pattern, whereas only angular constraints without perturbation of the model compound were used in the present work. These methodology differences may explain part of the differences between  $\Delta E_{\text{IDEAL}}$  and our new data. In addition, some differences can be explained by using nonideal starting structures. For instance, our stabilization energies for the C-9BPh family are significantly lower than those reported by Zirbel et al.<sup>24</sup> In this case, the  $\text{CH}\cdots\text{phosphate}$  interaction typical for the C-9BPh family is very weak and there is a possibility for the phosphate to create another, much stronger interaction with the neighboring  $\text{NH}_2$  group of cytosine. The latter interaction, although not fully developed within our constraints, contributes significantly to stabilization of both C-9BPh/C47-C366/1J5E and C-9BPh/C110-A109/1J5E structures. It is possible that the C-9BPh family exists as a local minimum in the X-ray structures as a consequence of the structural constraints imposed by the RNA structures.

The stability of (solvent-corrected) BPh interactions is in the range 0 to  $-11$  kcal/mol. Thus, although the solvent attenuates the absolute stability differences, it is clear that the BPh interactions show very variable energetics. The strongest BPh interactions appear to be stronger than the canonical AT base pair (the interaction energies of base pairs calculated by the same procedure are  $-9.3$ ,  $-14.7$ , and  $-18.6$  kcal/mol for AT, GC, and  $\text{CCH}^+$  base pairs, respectively). Note, however, that this stability assessment does not take into consideration structural effects, i.e., compatibility of the shapes of the studied interactions with the overall RNA structure. These factors will provide context-dependent tuning of the stability of the individual instances of BPh interactions. In other words, as is common in nucleic acids, a given interaction can have different effects on the stability of nucleic acids when occurring in different structural contexts, depending on its detailed context-specific interplay with the surrounding interactions.<sup>81</sup> It is not possible to unambiguously derive the amount of stabilization of nucleic acids structures associated with the individual interactions using model computations of interaction energies.<sup>1</sup>

The last four columns show DFT-D,<sup>51</sup> DFT-D3,<sup>73</sup> SAPT, and force field (MM) data. These calculations do not include the solvation term and thus should be compared with the fourth column, i.e., the CBS(T) gas phase interaction energies. The DFT-D and DFT-D3 calculations agree nicely with the reference data with average signed errors  $-0.2$  and  $-1.2$  kcal/mol and rmsd 0.6 and 0.7 kcal/mol, respectively. In fact, the older DFT-D variant (developed by one of us) seems to perform slightly better than the newer DFT-D3 method for the BPh interactions (mainly for the strongly bound complexes, where DFT-D3 slightly overbinds). DFT + dispersion based methods belong to the most successful ones for intermolecular forces. It is interesting that the DFT-D3 method, at least the variant we have used, systematically overbinds the BPh interactions between the nucleobases and an anionic phosphate groups. The SAPT calculations basically follow the CBS(T) trends but typically modestly underestimate the binding. The difference is likely due to smaller basis set used in the SAPT calculations.

The MM force field rather significantly underestimates the strength of the BPh interactions. The difference compared to CBS(T) data is in the range 2–10 kcal/mol. Basically, there are several obvious reasons why the MM description underestimates the BPh binding and why the underestimation is so nonuniform. One reason is the lack of polarization that is only partially compensated for by the HF-derived MM charges overestimating the static polarity of the interacting monomers. This affects basically all geometries. Another reason, also documented in some of our earlier studies for some other types of interactions, is overestimation of short-range repulsion effects by the Lennard-Jones term of the force field.<sup>82–84</sup> The  $r^{-12}$  repulsive term of the force field is too steep. This contributes to the QM vs MM difference in cases where there is a close separation of the monomers. Finally, we adapted fixed-point charges taken from the Cornell et al.<sup>76</sup> AMBER force field in our calculations as we primarily focused on assessment of the performance of the currently used force field. The fixed-point charges might poorly represent variable structures of the backbone segments.

In principle, the agreement between the QM and MM calculations could be improved by using charges fit for each conformation separately. Thus, we have done few such computations (see Table S3 in Supporting Information). Surprisingly, even in this case the difference between MM and QM calculations are quite nonuniform. One of the reasons could be too small size of the model systems, which might limit the accuracy of the atom-centered point charge model. This would concern mainly the anionic and flexible backbone segment, because in vacuo comparisons of QM and MM data for stacking and pairing interactions between planar nucleobases show somewhat smaller variability of the differences.<sup>85</sup> Further analysis of this problem is beyond the scope of the paper and will be done in future. It needs to be underlined that force fields are intended for condensed phase simulations of large biomolecular systems. Their benchmarking for small systems in the gas phase is very important,<sup>83,86</sup> but it is not the ultimate test to assess the force field performance. The QM literature sometimes tends to exaggerate the force field errors.

Nevertheless, Table 1 shows that for the strongest BPh interactions (more stable than  $-20$  kcal/mol in gas phase) in most cases the force field still provides relatively good description with an error smaller than 5 kcal/mol. The largest error for these strong interactions is seen for the C-9BPh/C47-C366/1J5E interaction and amounts to 8 kcal/mol. Therefore, for most

Table 2. Decomposition of the SAPT Gas Phase Interaction Energies (kcal/mol)<sup>a</sup>

	$E_{\text{el}}^{(1)}$	$E_{\text{exch}}^{(1)}$	$E_{\text{ind}}^{(2)}$	$E_{\text{disp}}^{(2)}$	$\delta(\text{HF})$	$E^{\text{SAPT}}$
A-2BPh/A315-G354/1JSE	-2.82	8.47	-4.53	-3.29	-1.21	-3.38
A-2BPh/A694-U788/1JSE	1.44	2.22	-1.99	-2.74	-0.11	-1.18
A-2BPh/A790-G1497/1JSE	-1.69	7.01	-3.73	-3.95	-0.74	-3.09
A-6BPh/A288-A119/1JSE	-14.32	14.94	-6.51	-5.76	-2.85	-14.50
A-6BPh/A353-G317/1JSE	-14.95	12.49	-6.12	-4.33	-2.41	-15.33
A-6BPh/A382-G64/1JSE	-16.06	14.40	-6.68	-4.36	-2.94	-15.64
A-7BPh/A886-C2114/1S72	-13.19	11.90	-6.06	-4.34	-2.14	-13.82
A-7BPh/A1233-C2526/1S72	-17.38	20.34	-7.37	-5.93	-3.95	-14.28
G-1BPh/G27-G297/1JSE	-27.81	23.35	-7.83	-7.10	-4.33	-23.71
G-1BPh/G394-U367/1JSE	-19.77	11.24	-4.84	-4.76	-1.46	-19.59
G-1BPh/G725-G666/1JSE	-28.20	26.19	-8.13	-7.50	-5.25	-22.88
G-3BPh/G111-A60/1JSE	-25.35	12.58	-6.84	-3.89	-2.74	-26.23
G-3BPh/G159-A162/1JSE	-32.47	27.28	-9.11	-7.26	-5.64	-27.20
G-3BPh/G31-C48/1JSE	-31.13	28.21	-8.76	-9.74	-5.17	-26.59
G-4BPh/G324-A327/1JSE	-33.99	36.01	-11.16	-10.81	-6.73	-26.68
G-4BPh/G326-G108/1JSE	-39.57	35.51	-12.79	-9.23	-7.02	-33.10
G-4BPh/G38-A397/1JSE	-29.50	24.70	-9.85	-8.17	-4.87	-27.69
G-5BPh→4BPh/G791-U1498/1JSE	-30.48	30.58	-10.46	-8.98	-5.80	-25.13
G-5BPh/G410-A431/1JSE	-31.27	21.01	-10.15	-6.59	-4.02	-31.01
G-5BPh/G800-A781/1JSE	-29.82	19.05	-9.57	-7.28	-3.84	-31.47
C-6BPh/C1127-G2070/1S72	-25.41	21.11	-7.19	-6.22	-3.17	-20.88
C-6BPh/C461-G479/1S72	-23.67	34.97	-9.80	-10.65	-6.23	-15.37
C-7BPh/C54-A353/1JSE	-27.86	21.85	-8.50	-6.18	-4.43	-25.12
C-7BPh→8BPh/C268-G266/1JSE	-25.32	20.30	-7.04	-7.95	-3.49	-23.50
C-8BPh/C1484-A1483/1JSE	-26.83	21.41	-8.19	-7.22	-3.58	-24.40
C-8BPh/C283-A282/1JSE	-27.99	22.15	-8.54	-6.78	-4.24	-25.40
C-9BPh/C47-C366/1JSE	-26.92	23.81	-6.28	-9.57	-3.09	-22.05
C-9BPh/C110-A109/1JSE	-19.57	13.48	-5.61	-7.14	-1.95	-20.80
U-5BPh/U14-U17/1JSE	-14.74	21.18	-8.24	-6.35	-4.67	-12.83
U-5BPh/U1506-U1522/1JSE	-18.04	28.35	-10.14	-7.06	-6.52	-13.41
U-9BPh/U1049-C1203/1JSE	-5.76	3.14	-3.37	-2.67	-0.47	-9.13
U-9BPh/U14-U13/1JSE	0.03	0.86	-2.63	-1.65	-0.09	-3.48

<sup>a</sup>  $E_{\text{el}}^{(1)}$  = electrostatic component,  $E_{\text{exch}}^{(1)}$  = exchange repulsion,  $E_{\text{ind}}^{(2)}$  = induction component,  $E_{\text{disp}}^{(2)}$  = dispersion component,  $\delta(\text{HF})$  = correction for higher-order induction terms.

BPh interactions we assume that the force field may allow satisfactory description in explicit solvent simulations, at least locally, when the simulation starts from correctly folded experimental geometry. It is not clear if the force field description of BPh interactions would be sufficiently accurate for the purpose of simulation studies attempting RNA folding and RNA structure prediction. In general, we do not assume that the present force fields are matured enough for such tasks.<sup>87</sup> When all the data are evaluated, it appears that due to the nonpolarizable nature of the force field model, the strength of the BPh interactions might be systematically underestimated in the gas phase by  $\sim 3$ – $5$  kcal/mol. This is a slightly better “in vacuo” accuracy than in the case of monovalent cation–nucleobase interactions (see below for further discussion).<sup>82</sup>

Table 2 summarizes the energy decomposition using the SAPT method. There are basically two types of interactions in the BPh families, hydrogen bonding and weak  $\text{CH}\cdots\text{phosphate}$  interactions. Regarding the SAPT interaction energy components, the hydrogen bonded motifs are very similar to the other common hydrogen bonds. The electrostatic component constitutes about 60% of the total attractive energy. Induction

(including the  $\delta\text{HF}$  term) and dispersion contribute about 25 and 15%, respectively. The only difference between regular hydrogen bonds and hydrogen bonds to the charged phosphate groups is perhaps the somewhat larger contribution of the induction energy in the latter (in regular hydrogen bonds induction contributes usually less than 20% of the total attraction). The larger induction energy is because the phosphate groups are charged. The weakly bound  $\text{CH}\cdots\text{phosphate}$  complexes follow significantly different interaction patterns and can be distinguished from the hydrogen bonded ones by smaller importance of the electrostatic interaction, which is usually smaller in magnitude than the induction component.

#### Base–Phosphate Interactions in the Hairpin Ribozyme.

Table 3 and Figure 3 show data for selected binding modes of BPh interactions occurring in the hairpin ribozyme (HrRz; see Supporting Information for 2D structure of HrRz, Figure S2). Four of the calculated structures (HrRz str4–str7) are unique as they involve N1-protonated adenine. Such BPh pattern has not been considered in the original classification.<sup>24</sup> The protonated adenine 38 is assumed to be involved directly in RNA self-cleavage,<sup>88,89</sup> and its likely role is in electrostatic stabilization of

the transition state and/or in general acid catalysis.<sup>42,78,90–93</sup> The MM force field performs quite well for this interaction. The gas phase error of  $\sim 10$ – $15$  kcal/mol ( $\sim 10$ – $15\%$  of total interaction gas phase energy) is rather acceptable. We suggest that the description in the course of explicit solvent simulation, which is assumed to be improved by compensation of errors to a significant extent compared to the in vacuo accuracy, may be rather meaningful. Note that in the course of MD simulation the absolute gas phase accuracy of the individual terms is not the most important issue. The key issue is relative binding strength of different groups with different partners including the competing water molecules. The relative error in the gas phase MM description of the BPh interaction with protonated nucleobase is comparable to or better than that for interactions between nucleobases and monovalent ions<sup>82,94</sup> and much better than for binding of divalent ions, such as  $\text{Mg}^{2+}$ .<sup>95</sup> Note that it is usually assumed that description of monovalent ion binding to nucleic acids in explicit solvent simulations is rather satisfactory, in contrast to binding of divalent ions where the MM description by pair-additive force field is already quite unrealistic.<sup>96</sup> It is thus

**Table 3. Total Interaction Energies (kcal/mol) for BPh Interactions with Model Structures Taken from MD Simulations of the Hairpin Ribozyme<sup>a</sup>**

	MP2/aDZ	$\Delta E_{\text{COSMO}}^{\text{MP2/aDZ}}$	MM
HrRz str1	−1.41	1.76	1.05
HrRz str2	−13.08	−2.19	−10.29
HrRz str3	−16.71	−4.13	−11.48
HrRz str4	−108.38	−16.16	−91.71
HrRz str5	−110.52	−16.60	−88.16
HrRz str6	−105.93	−13.71	−92.24
HrRz str7	−85.82	−6.64	−75.42
HrRz str8	−16.75	−5.40	−10.10
HrRz str9	−8.45	−1.89	−5.29
HrRz str10	−8.94	−1.42	−4.82

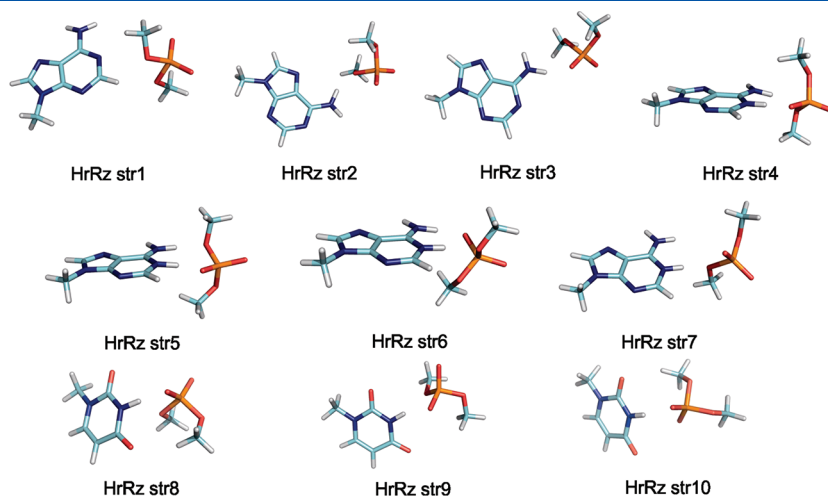
<sup>a</sup> The protonation state of A38 in real structures is not known and likely changes when the catalysis proceeds. The simulations done with canonical and protonated A38 sample different local substates and were used to generate a broad range of structures for QM analysis (Figure 3).<sup>42</sup>

fair to assume that accuracy in description of BPh interactions by pair-additive force fields in simulations may be comparable to or better than the accuracy of description of monovalent ion binding to the solute binding sites, although probably less realistic than description of base pairing and stacking.

The molecular dynamics simulations suggested that the most populated BPh pattern among all intrinsic conformations of A38 and scissile phosphate (structures HrRz str2–str7, the structure HrRz str1 corresponds to X-ray geometry) is the HrRz str6 corresponding to the 4BPh interaction of protonated A38H<sup>+</sup> with the scissile phosphate (Table 4). In this interaction pattern, the A38H<sup>+</sup>(N1) nitrogen donates the H-bond to the G+1(O5') oxygen, which becomes the leaving alcoholate during the self-cleavage reaction. This observation together with high interaction energy between protonated adenine and phosphate within 4BPh motif supports the role of protonated adenine A38H<sup>+</sup> in electrostatic stabilization of the transition state and/or in general acid catalysis. Note that the actual protonation state and binding pattern of A38 in HrRz so far could not be unambiguously determined by experiments and are likely variable, with A38 pK<sub>a</sub> shift toward neutrality. Therefore, we cannot assess quality of the force field using these simulations as benchmarks.<sup>42</sup>

**Base–Phosphate Interactions in the Sarcin–Ricin Loop Motif.** The sarcin–ricin loop motif (SRL) is one of the most important recurrent RNA motifs, i.e., molecular building blocks shaped by specific intermolecular interactions with characteristic sequence signature (see ref 30 and references therein). Originally, the SRL motif was found in the highly conserved sarcin–ricin domain (SRD) of 23S–28S rRNAs from the large ribosomal subunit. The SRL motif forms a crucial site for stimulation of GTP-ase activity of elongation factors during translation. The SRD domain is shaped by several extended non-Watson–Crick base pairs bolstered by several BPh interactions. The SRD domain is stiff and behaves very stably in simulations. Thus, we have derived structures of key BPh interactions for QM computations from the simulation trajectories; the geometries agree well with the X-ray data. The results are shown in Figure 4 and Table 5. Figure 4 illustrates how the BPh interactions complement the RNA base pairing.

The SRD is composed of three structural segments: (i) the GNRA tetraloop, (ii) the G-bulge region containing bulged guanine, which is the most critical nucleotide for elongation



**Figure 3.** Structures of the calculated BPh interactions taken from the MD simulations of the hairpin ribozyme.

factor G binding, and (iii) the flexible region (see Supporting Information for structure of SRD, Figure S2).<sup>30</sup> The RNA strand between G-bulged and flexible regions forms a specific S-turn shape. In this backbone path, the adenine upstream to the bulged

**Table 4.** Dynamical Stability of the BPh Interactions of the Hairpin Ribozyme in the 50-ns-Long and 150-ns-Long Molecular Dynamics Simulations Having Canonical and Protonated Form of Catalytic Adenine A38, Respectively, Represented by Occupancy of the BPh Hydrogen Bonds in %, As Defined in the Methods

structure <sup>a</sup>	H-bonds	occupancy (%)
Str2	A38(N6)⋯G+1(O5')	4.9
Str3	A38(N6)⋯G+1(O1P)	1.0
Str4	A38H <sup>+</sup> (N1)⋯G+1(O1P)	1.1
	A38H <sup>+</sup> (N6)⋯G+1(O5')	9.7
Str5	A38H <sup>+</sup> (N6)⋯A-1(O3')	9.3
	A38H <sup>+</sup> (N1)⋯G+1(O1P)	1.1
	A38H <sup>+</sup> (N1)⋯A-1(O3')	17.0
Str6	A38H <sup>+</sup> (N6)⋯G+1(O2P)	68.5
	A38H <sup>+</sup> (N1)⋯A-1(O3')	17.0
	A38H <sup>+</sup> (N1)⋯G+1(O2P)	26.4
	A38H <sup>+</sup> (N1)⋯G+1(O5')	76.9
Str7	A38H <sup>+</sup> (N1)⋯G+1(O5')	76.9
Str8	U37(N3)⋯U+2(O1P)	61.0
Str9	U37(N3)⋯G+1(O3')	12.8
Str10	U37(N3)⋯U+2(O2P)	0.0

<sup>a</sup> The population of BPh interactions of structures Str2 and Str3 are calculated from WT1/G8/A38 simulation, and others are calculated from WT2/G8/A38H<sup>+</sup> simulation; these simulations are reported in ref 42.

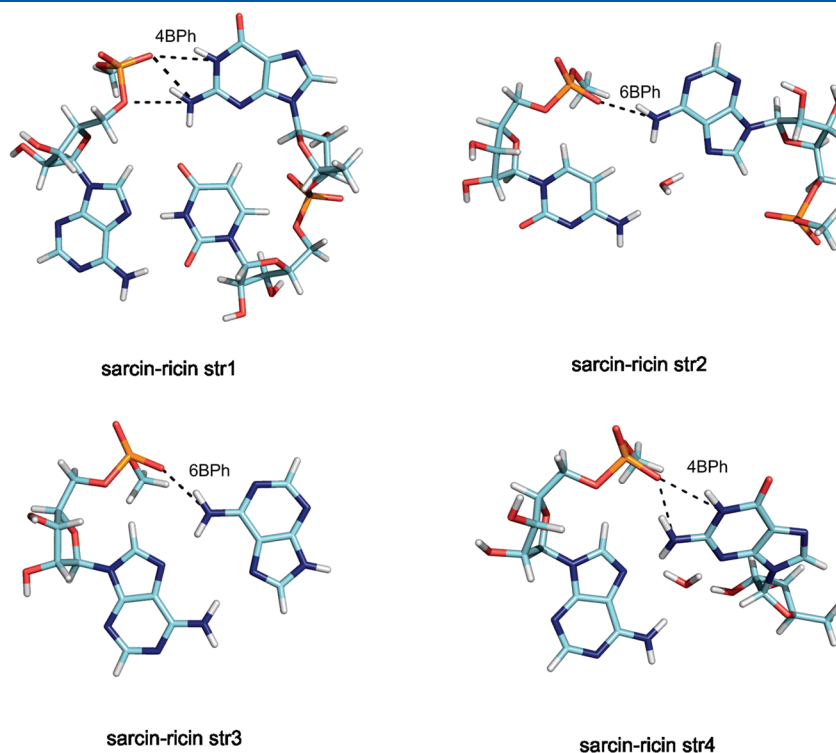
guanine is flipped over and stacked to the bulged guanine by opposite stacking face in comparison with stacking in canonical A-form RNA. The next upstream nucleotide is flipped back to canonical conformation. These flips form two sharp bends of the sugar–phosphate backbone giving it the S-shape appearance.<sup>29</sup>

The S-turn region contains three closely spaced BPh interactions, of which two are analyzed in this study. The bulged guanine G2655 (numbering according *Escherichia coli* 23S rRNA) forms a 4BPh with the phosphate of adenine A2665. Both nucleotides participate in the base triplet of the G-bulged region (structure str1). The A2654 upstream to the bulged guanine forms 6BPh

**Table 5.** Total Interaction Energies (kcal/mol) for BPh Interactions (See Figure 4 for the Fragments Used in the Computations) Taken from MD Simulations of Sarcin–Ricin Internal Loops<sup>30</sup> of *E. coli* and Rat<sup>a</sup>

	MP2/aDZ	$\Delta E_{\text{COSMO}}^{\text{MP2/aDZ}}$	MM
sarcin–ricin str1	−5.27	−14.79 <sup>b</sup>	−4.68
sarcin–ricin str2	−5.69	−11.74 <sup>b</sup>	−4.78
sarcin–ricin str3	−24.63	−10.72	−19.16
sarcin–ricin str4	−58.57	−21.66	−61.13
sarcin–ricin str2 ww	12.04	−2.73 <sup>b</sup>	12.46
sarcin–ricin str4 ww	−34.47	−9.43	−37.33

<sup>a</sup> The simulated sarcin–ricin loops are very stable and thus provide examples of important structural BPh interactions.<sup>30</sup> Interaction energies for structures 2 and 4 were calculated as trimers, i.e., including water as the third molecule. “ww” means that the water molecule was excluded from the interaction energy computation. <sup>b</sup> Interaction energies for structures 1 and 2 are more stabilizing in the COSMO water environment because COSMO water screens out the repulsive long-range electrostatic interaction between two negatively charged phosphates in these particular systems.



**Figure 4.** Structures of the sarcin–ricin loop BPh interactions.



**Table 6.** Dynamical Stability of the BPh Interactions (and Some Other Proximal Interactions) of the Sarcin–Ricin Internal Loop Motif from *E. coli* and Rat 23S and 28S rRNAs in 50-ns-Long Molecular Dynamics Simulations Represented by Occupancy of the BPh Hydrogen Bonds in %<sup>a</sup>

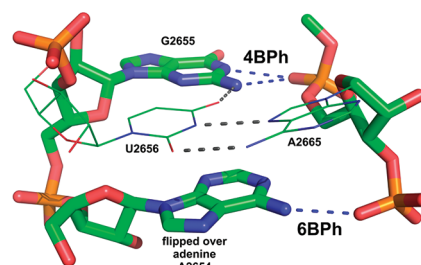
structure	H-bonds	occupancy (%)	
		rat	<i>E. coli</i>
Str1	G2655(N1)⋯A2665( <i>pro</i> -R <sub>p</sub> )	98.0	92.7
	G2655(N2)⋯A2665(OS')	81.5	47.2
	G2655(N2)⋯A2665( <i>pro</i> -R <sub>p</sub> )	0.0	95.5
Str2	A2654(N6)⋯C2666( <i>pro</i> -R <sub>p</sub> )		96.2
	A2654(N7)⋯C2666(N4)		11.7
	A2654(N7)⋯w⋯C2666(N4)		13.7
Str3	A4318(N6)⋯A4330( <i>pro</i> -R <sub>p</sub> )	97.8	
	A4318(N6)⋯A4330(N7)	99.7	
	A4318(N7)⋯A4330(N6)	99.3	
Str4	G2659(N2)⋯A2662( <i>pro</i> -R <sub>p</sub> )	86.0	99.6
	G2659(N1)⋯A2662( <i>pro</i> -R <sub>p</sub> )	81.4	93.0
	G2659(N3)⋯A2662(N6)	3.4	0.0
	G2659(N3)⋯w⋯A2662(N6)	17.9	18.8

<sup>a</sup> The numbering is according to the *E. coli* structure except for the structure str3, which is numbered according to rat. For more details about sarcin ricin loop simulations see ref 30.

interaction with C2666, to which it is also base paired (structure str2 in *E. coli* SRD and str3 in rat SRD). Consistently with calculated large (in absolute values) interaction energies of these motifs, both these BPh interactions were highly populated in molecular dynamics (Table 6). It is interesting that in *E. coli* SRD, the 6BPh interaction between A2654 and phosphate of C2666 was even more populated than the H-bonds forming the tHH (trans Hoogsteen–Hoogsteen) A2654/C2666 base pair, suggesting that the 6BPh interaction between A2654 and C2666 dominates the overall stabilization of the tHH A/C pairing region (Table 6). In addition, the directionality of both above-noted BPh interactions most likely determines the flipped over conformation of A2654 with respect to G2655, and thus the first sharp bend in the S-turn (Figure 5). We suggest that the BPh interactions of G2655 and A2654 stabilize the flipped over conformation of the adenine A2654 and are crucial for shaping the sugar–phosphate backbone into the S-turn path, most likely being even more important than the specific base pairing itself. This is supported by the conservation of A2654 and G2655, whereas there is no evolutionary pressure to the 2666 nucleotide paired to A2654. In other words, the nucleotide at position 2654 should be adenine in flipped over conformation to be able to form the 6BPh interaction with phosphate from the second strand, whereas the H-bonds between the 2654 and 2666 nucleobases are variable and not conserved.

The last structure str4 corresponds to the tHS A/G base pair from the pedestal of the GNRA tetraloop. The tHS A/G base pair is an example of water-mediated base pairing that is stabilized in simulations by water molecules with mean residency time of ~300 ps. The proximal 4BPh interaction was highly stable in the simulations and likely represents one of the dominant interactions in this region (Table 6).

BPh interactions in SRL are very well determined from high-resolution X-ray structures and the simulations thus indicate



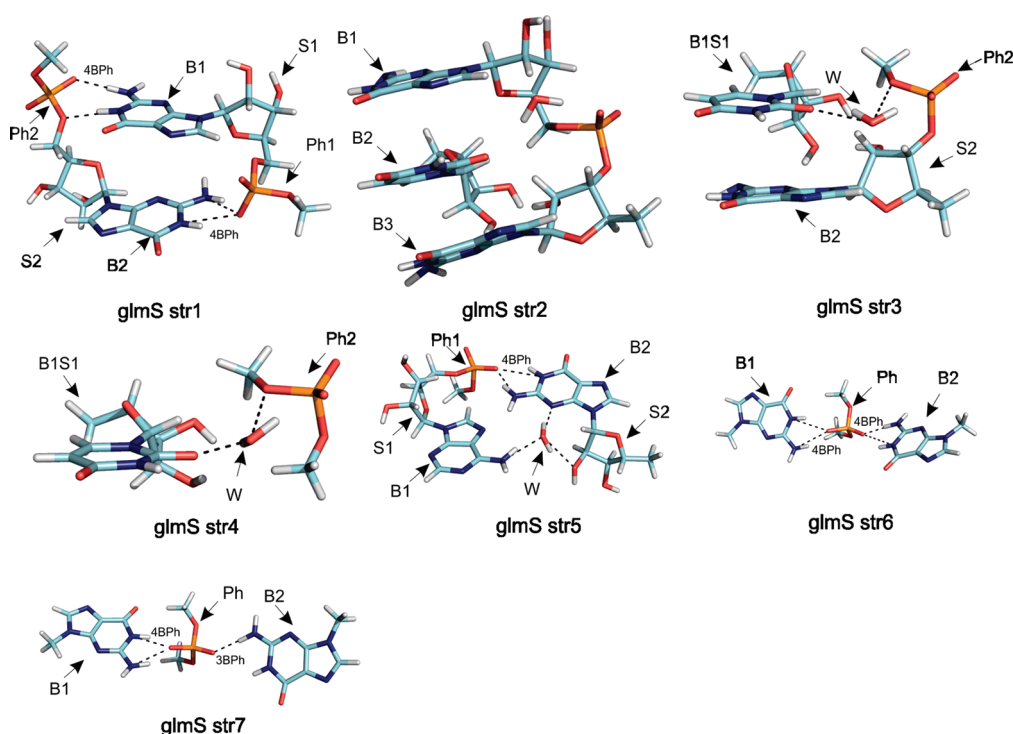
**Figure 5.** First bend of S-turn submotif from *E. coli* 23S rRNA SRD showing flipped over adenine A2654 and stabilization of this topology by BPh interactions. The pairing interactions within base triple of G-bulge region are shown in gray, and BPh H-bonds are in blue.<sup>24,29,30</sup>

good performance of the force field for BPh interactions in simulations.

**Complex Base–Phosphate Interactions in the *glmS* Riboswitch.** The *glmS* riboswitch-ribozyme is part of the 5'-untranslated region of mRNA coding glucosamine-6-phosphate synthetase in numerous Gram-positive bacteria. It can also be classified as a ribozyme as binding of glucosamine-6-phosphate results in catalytic self-cleavage of the sugar–phosphate backbone followed by degradation of mRNA ultimately resulting in down-regulation of glucosamine-6-phosphate synthetase expression.<sup>97–100</sup> Almost every noncanonical region of *glmS* riboswitch is accompanied by at least one BPh interaction (see Supporting Information for the 2D structure of *glmS* riboswitch, Figure S2).<sup>41</sup> This fact underlines that the BPh interactions often participate in complex networks of highly specific and conserved molecular interactions, which is not apparent from the basic BPh classification.<sup>24</sup>

We analyzed BPh interactions taken from four different parts of the *glmS* riboswitch. The most important BPh contacts are located in the active site of the *glmS* riboswitch between scissile phosphate and guanines G39 and G65, respectively (structures *glmS* str6 and str 7; see Figure 6). These BPh interactions are important for self-cleavage reaction as they stabilize reactive conformation of the scissile phosphate and increase the reactivity of phosphate by polarization of its electron density. Table 7 documents that these BPh contacts are the most stable BPh interactions in the *glmS* riboswitch in molecular dynamics simulations. The rigidity of the active site and high stability of these two BPh interactions support a hypothesis about their structural and electrostatic stabilization function in self-cleavage catalysis. The native geometry of this segment is represented by structure *glmS* str7 having 3BPh interaction between guanine G65 (here labeled as base B2) and scissile phosphate. This BPh contact was, however, relaxed to 4BPh interaction (structure *glmS* str6) during geometry minimization of our model QM system. The 3BPh interaction in the native structural context is stabilized by additional H-bond contact of G65(N1) nitrogen with the hydroxyl group of glucosamine-6-phosphate ligand, which in turn makes another H-bond to the nonbridging oxygen of scissile phosphate.<sup>41</sup> This interaction is not included in the QM model system. Thus although 3BPh or 5BPh interactions can easily relax to the energetically more favorable 4BPh contact, they might still be preferred when they are accompanied by other interactions in their full structural context.

The structure *glmS* str5 originates from the tHS A/G base pair of the GNRA tetraloop and is thus similar to the corresponding SRD structure str4, which is also taken from a GNRA tetraloop.



**Figure 6.** Structures of *glmS* riboswitch-ribozyme BPh interactions with indicated fragment names. See also the text for further explanation of the location of the interactions in the folded ribozyme.<sup>41</sup> Note that str3 and str4 are fragments subtracted from str2, shown in addition from different directions (the depicted water molecule can be used for orientation). The numbering is entirely arbitrary, reflecting the individual monomers used in QM computations (it differs for str2 and str3–4) and not the nucleotide numbering.

**Table 7. Stability (Occupancy in %) of Selected BPh Interactions (and Some Other Proximal Interactions, Figure 6) of the *glmS* Riboswitch from *Thermotoga maritima* in Molecular Dynamics Simulations<sup>a</sup>**

structure	H-bonds	population (%)
Str1	G94(N1)···G138(O5')	43.0
	G94(N1)···G138( <i>pro-S<sub>p</sub></i> )	76.0
	G94(N2)···G138( <i>pro-S<sub>p</sub></i> )	93.9
	G94( <i>pro-R<sub>p</sub></i> )···G138(N1)	79.9
	G94( <i>pro-R<sub>p</sub></i> )···G138(N2)	81.2
Str2–4	U67(O2)···w···G41(O5')	22.2
	U67(O2)···w···G41( <i>pro-R<sub>p</sub></i> )	21.4
Str5	G114(N1)···A117( <i>pro-R<sub>p</sub></i> )	98.1
	G114(N2)···A117( <i>pro-R<sub>p</sub></i> )	100
	G114(N3)···A117(N6)	0.8
	G114(N3)···w···A117(N6)	31.0
Str6–7	G39(N1)···G1( <i>pro-S<sub>p</sub></i> )	100
	G39(N2)···G1( <i>pro-S<sub>p</sub></i> )	99.8
	G65(N2)···G1( <i>pro-R<sub>p</sub></i> )	96.9

<sup>a</sup> The simulations model different protonation states of the bound ligand with the canonical form of G40.<sup>41</sup> The Table shows examples of versatile interactions that can contribute to formation of the catalytic core as well as stable interactions important for folding of the RNA. The BPh interactions except for the structure Str1 were evaluated on the basis of 50-ns-long G40/GlcN<sup>6</sup>P<sup>2−</sup> simulation with the dominant glucosamine-6-phosphate ligand protonation state. The BPh populations of structure Str1 were calculated from 20-ns-long G40/GlcN<sup>6</sup>P<sup>2−</sup> simulation with the minor protonation state of the ligand, as this arrangement was not populated in simulations with the dominant ligand protonation form.<sup>41</sup>

Similarly to SRD, the *tHS* A/G base pair forms a water mediated base pair (the residency times of water molecules were ~500 ps) with highly stable 4BPh interaction (Table 7).

The structure *glmS* str1 is an interesting motif, where two nucleotides are stacked on each other, so that the base of the first nucleotide makes a BPh contact to the phosphate of the second nucleotide and vice versa. This geometry is, however, located in a flexible part of *glmS* riboswitch and was observed only in MD simulations but not in X-ray. Furthermore, this interaction was established in equilibration of only one of seven molecular dynamics simulations (the G40/GlcN<sup>6</sup>P<sup>2−</sup> simulation; see ref 41 for the details), where it was stably populated but was not sampled in the rest of the simulations, so its relevancy is debatable.

The remaining structures *glmS* str2–4 correspond to the last fragment of the *glmS* riboswitch. The structures str3 and str4 are fragments of more complete structure str2, which include a water-mediated interaction between base and phosphate rather than a true BPh contact. The water involved in these structures was found to be a long residency water molecule in our recent molecular dynamics simulations with residency time of ~500 ps (Table 7). This water mediated interaction between base and phosphate was crucial for *glmS* active site arrangement as any structural changes of this region resulted in distortion of the *glmS* active site.<sup>41</sup> The interaction energies show that this water-mediated BPh bridge, once formed in the structural context, could be as strong as the direct BPh interaction (Table 8). Note that water-mediated BPh interactions are not included in the current BPh interaction classification.<sup>24</sup> We suggest that extension of the classification to include water-mediated interactions would be useful and that MD simulations may represent a plausible tool to study such interactions.

**Table 8. Total Interaction Energies (kcal/mol) of BPh Interactions (and Some Other Proximal Interactions) Taken from the *glmS* Riboswitch (Figure 6)<sup>a</sup>**

	fragment	MP2/aDZ	$\Delta E_{\text{COSMO}}^{\text{MP2/aDZ}}$	MM
str1-frag1	B1...B2	−2.21	−1.87	−2.45
str1-frag2	B1...Ph2	−31.77	−9.54	−28.54
str1-frag3	B2...Ph1	−28.58	−7.38	−30.35
str1-frag4	B1S1Ph1...B2	−34.61	−12.01	−34.00
str1-frag5	B2S2Ph2...B1	−29.88	−11.04	−34.94
str1-frag6	B1S1Ph1...B2S2Ph2	−26.90	−20.96	−37.27
str2-frag1	B1...B2	−7.87	−3.93	−6.18
str2-frag2	B2...B3	−7.37	−4.97	−7.45
str3-frag1	B1S1...W...B2S2Ph2	−34.69	−19.57	−38.09
str3-frag2	B1S1+B2S2Ph2...W	−16.96	−6.22	−17.34
str3-frag3	B1S1...W+B2S2Ph2	−20.84	−15.61	−24.55
str3-frag4	B1S1+W...B2S2Ph2	−31.07	−18.25	−34.28
str4-frag1	B1S1...Ph2...W	−13.44	−1.26	−18.86
str5-frag1	B1S1Ph1...B2	−33.36	−9.60	−34.85
str5-frag2	B1S1Ph1+B2S2...W	−16.26	−8.32	−17.04
str5-frag3	B1S1Ph1...B2S2+W	−43.34	−14.07	−45.31
str6-frag1	B1...Ph...B2	−55.45	−15.14	−52.93
str6-frag2	B1...Ph	−29.81	−9.14	−26.87
str6-frag3	B2...Ph	−28.37	−6.23	−26.36
str6-frag4	B1...B2	0.21	−0.03	0.31
str7-frag1	B1...Ph...B2	−55.59	−14.35	
str7-frag2	B1...Ph	−32.74	−9.75	
str7-frag3	B2...Ph	−25.64	−4.77	
str7-frag4	B1...B2	0.29	−0.03	

<sup>a</sup> B, nucleobase; Ph, dimethylphosphate; S, sugar; W, water. See Figure 6. The “...” mark divides the whole system into the two or more subsystems taken in the interaction energy computations formally as the monomers.

In summary, the BPh interactions behave usually very stably in explicit solvent MD simulations (Table 7). The comparison of MM interaction energies with reference QM values documents that BPh interactions are reasonably described at MM level. The maximum error in gas phase interaction energies of BPh interactions selected by us (for fragments containing only base and phosphate) is less than 3 kcal/mol (Table 8). There are larger errors of MM gas phase energies (−10.4 kcal/mol in case of fragment str1-frag6 from Table 8) in some cases. However, they may be caused also by additional terms, such as differences in stacking energies etc. As we already noted, for appropriate performance of the simulations the overall relative balance of different terms is crucial, so the errors in the gas phase MM description of the BPh interactions are likely acceptable, as also evidenced by the simulation behavior (Tables 4, 6, and 7).

## CONCLUSIONS

Complex shapes of functional RNA molecules are primarily determined by an astonishing variability of noncanonical interactions. Base–phosphate (BPh) H-bonds belong to the strongest nonbonded interactions found in RNA. Despite this fact, they seemed to be mostly overlooked in the past. The importance of BPh interactions was pointed out only recently by Zirbel et al.,<sup>24</sup> who also presented their systematic classification and revealed their evolutionary roles. It turned out that ~12% of

nucleotides in large RNAs are involved in direct BPh interactions. These interactions in addition belong to the most conserved during evolution.

In the present work we extend the above-mentioned study by performing a series of calculations including QM benchmark calculations on 32 individual BPh structures, DFT calculations, SAPT decompositions, and force field (AMBER Cornell et al.<sup>76</sup>) calculations. We in addition analyze a representative set of larger models of BPh interactions taken from specific noncoding RNAs, namely ribosomal sarcin–ricin loop domain, hairpin ribozyme, and *glmS* riboswitch-ribozyme. Besides the structure-energy computations, a representative set of explicit solvent simulation trajectories is analyzed, to capture the structural stability of BPh interactions in the course of simulations. Our effort is aimed at gaining additional physical chemistry insight into this interesting and highly relevant type of H-bond.

The reference CBS(T) computations reveal a wide range of interaction energies of the individual instances of the BPh interactions, −1 to −36 kcal/mol in gas phase and 0 to −11 kcal/mol upon adding the continuum solvent screening. Although the reference computations in most cases agree well with the earlier classification, in several instances we observed rearrangement of the interaction to a neighboring family, or formation of a doubly bonded (bifurcated) structure that has not been considered in the original classification.<sup>24</sup>

The SAPT decomposition analysis indicates that the base–phosphate interactions involving polar hydrogens are mainly of electrostatic origin, with marked induction component and also a significant dispersion component. Therefore, although the BPh interactions are significantly stronger than, e.g., the base–base hydrogen bonds (assuming the interaction energy normalized to one hydrogen bond), they are of quite similar origin, with somewhat larger induction contribution. On the other hand, the BPh interactions involving nonpolar carbon-bound hydrogen (C–H...phosphate) are much less electrostatic in nature, with pronounced or dominant dispersion and induction stabilization. These interactions are also much weaker and expected to play only minor role in structure stabilization of RNA.

Comparison with benchmark CBS(T) calculations points to the necessity of using large basis sets when accurate MP2 interaction energies are to be obtained. The CCSD(T) correction for higher-order correlation effects is less important, consistent with other studies on hydrogen bonding.<sup>79,101,102</sup> Very good interaction energies may be obtained also using much less demanding DFT-D (TPSS/6-311++G(3df,3pd)/0.96–27) method. SAPT interaction energies are on average somewhat underestimated, probably due to smaller basis set used.

To test the reliability of the frequently used Cornell et al.<sup>76</sup> force field, we calculated also the force field interaction energies in vacuo. The overall agreement with the reference QM calculations is quite good, although the interaction energies are in absolute value underestimated by the pair-additive force field. Note, however, that the in vacuo QM interaction energies are not directly comparable to the in vacuo force field energies, because the force field charges are intended for use in the condensed phase. Therefore, the errors in the condensed phase simulations will partially cancel out as compared to the in vacuo calculation. In addition, the performance of the force field in simulations depends mainly on mutual balance of various interactions, which may be satisfactory even with some errors in absolute gas phase energies. By comparing with other types of interactions, we suggest that the accuracy of the MM force field in gas phase



calculations is likely sufficient to obtain a qualitatively correct description of BPh interactions in simulations. This is subsequently confirmed by a set of explicit solvent simulations.

Some BPh interactions extracted from the X-ray structures were found not to be a genuine local minimum in QM optimization. They tended to switch to a neighboring, more stable family (see the reclassified structures in Figure 2). This is typical, e.g., for G-5BPh and G-3BPh families, which tended to relax into the more stable G-4BPh interaction. This indicates that they emerge as a well-defined family in structural analyses due to the context of larger RNA structures. This can be demonstrated, e.g., on G-3BPh interaction of guanine G65 in the active site of *glmS* riboswitch, which is constrained to be the 3BPh by the additional interaction of G65(N1) with the ligand in the active site of the riboswitch.

The examples of BPh interactions from three folded RNA structures (hairpin ribozyme, *glmS* riboswitch, and sarcin–ricin domain) show that the noncanonical parts of RNA structures are often accompanied by BPh interaction. The BPh interactions are often involved in intricate networks of additional interactions and play a crucial role in the stabilization of these noncanonical (and biologically very important) parts of RNA structures. This is most apparent in the case of the sarcin–ricin motif, in which the structure of the intricate S-turn submotif is most likely primarily determined by BPh interactions, with subsequent adaptation of the noncanonical base pairing. Another important role of BPh interaction suggested by our computations is their likely participation in RNA catalysis. Both examples of RNA enzymes have the scissile phosphate bound to the active site nucleotides by BPh interaction. These interactions thus likely play an important role in structural stabilization of the scissile phosphate and they participate in electrostatic stabilization of the transition state as they pull the electron density out of the scissile phosphate during nucleophile attack. Finally, we show that water-mediated BPh interactions, which could not be included in the original bioinformatics classification, may play structural and functional roles.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Detailed description of the constraints during optimization; structures with atom-naming conventions; deformation energies for a representative set of structures; MM interaction energies calculated with RESP charges fitted for a given conformation; the secondary structure of HrrZ, sarcin–ricin loop domain, *glmS* riboswitch; Cartesian coordinates of structures utilized in reference QM interaction energy computations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [sponer@ncbr.chemi.muni.cz](mailto:sponer@ncbr.chemi.muni.cz).

## ■ ACKNOWLEDGMENT

This work was supported by the Academy of Sciences of the Czech Republic [grant numbers AVOZ50040507 and AVOZ50040702], by the Ministry of Education of the Czech Republic [grant number LC06030], by the Grant Agency of the Academy of Sciences of the Czech Republic [grant number IAA400040802], by Student Project PrF\_2011\_020 of Palacky

University, and Grant Agency of the Czech Republic [grant numbers P208/10/2302, 203/09/1476, P208/11/1822, P208/10/1742, P301/11/P558, and 203/09/H046]. This work was supported by the project “CEITEC - Central European Institute of Technology” (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund. M.Z., P.B., P.J., and M.O. gratefully acknowledge the support by the Operational Program Research and Development for Innovations - European Regional Development Fund (projects CZ.1.05/2.1.00/03.0058 and CZ.1.07/2.3.00/20.0017 of the Ministry of Education, Youth and Sports of the Czech Republic).

## ■ REFERENCES

- (1) Sponer, J.; Sponer, J. E.; Petrov, A. I.; Leontis, N. B. *J. Phys. Chem. B* **2010**, *114*, 15723–15741.
- (2) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905–920.
- (3) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327–339.
- (4) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4899–4903.
- (5) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497–3531.
- (6) Mokdad, A.; Krasovska, M. V.; Sponer, J.; Leontis, N. B. *Nucleic Acids Res.* **2006**, *34*, 1326–1341.
- (7) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. *J. Math. Biol.* **2008**, *56*, 215–252.
- (8) Sponer, J.; Mokdad, A.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Leontis, N. B. *J. Mol. Biol.* **2003**, *330*, 967–978.
- (9) Sharma, P.; Sponer, J. E.; Sponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A. *J. Phys. Chem. B* **2010**, *114*, 3307–3320.
- (10) Sponer, J.; Zgarbova, M.; Jurecka, P.; Riley, K. E.; Sponer, J. E.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *9*, 1166–1179.
- (11) Sponer, J. E.; Leszczynski, J.; Sychrovsky, V.; Sponer, J. *J. Phys. Chem. B* **2005**, *109*, 18680–18689.
- (12) Sponer, J. E.; Reblova, K.; Mokdad, A.; Sychrovsky, V.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2007**, *111*, 9153–9164.
- (13) Sponer, J. E.; Spackova, N.; Kulhanek, P.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. A* **2005**, *109*, 2292–2301.
- (14) Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2005**, *109*, 11399–11410.
- (15) Oliva, R.; Tramontano, A.; Cavallo, L. *RNA* **2007**, *13*, 1427–1436.
- (16) Sharma, P.; Chawla, M.; Sharma, S.; Mitra, A. *RNA* **2010**, *16*, 942–957.
- (17) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2008**, *25*, 709–732.
- (18) Sharma, P.; Sharma, S.; Chawla, M.; Mitra, A. *J. Mol. Model.* **2009**, *15*, 633–649.
- (19) Oliva, R.; Cavallo, L.; Tramontano, A. *Nucleic Acids Res.* **2006**, *34*, 865–879.
- (20) Petrov, A. S.; Bowman, J. C.; Harvey, S. C.; Williams, L. D. *RNA* **2011**, *17*, 291–297.
- (21) Czyznikowska, Z.; Gora, R. W.; Zalesny, R.; Lipkowski, P.; Jarzemska, K. N.; Dominiak, P. M.; Leszczynski, J. *J. Phys. Chem. B* **2010**, *114*, 9629–9644.
- (22) Czyznikowska, Z.; Lipkowski, P.; Gora, R. W.; Zalesny, R.; Cheng, A. C. *J. Phys. Chem. B* **2009**, *113*, 11511–11520.
- (23) Toczyłowski, R. R.; Cybulski, S. M. *J. Phys. Chem. A* **2003**, *107*, 418–426.
- (24) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. *Nucleic Acids Res.* **2009**, *37*, 4898–4918.
- (25) Ulyanov, N. B.; James, T. L. *New J. Chem.* **2010**, *34*, 910–917.
- (26) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; Daubentoncarafa, Y.; Uhlenbeck, O. C.; Tinoco, I.



- Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1364–1368.
- (27) Hsiao, C.; Mohan, S.; Hershkowitz, E.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2006**, *34*, 1481–1491.
- (28) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3836–3849.
- (29) Correll, C. C.; Munishkin, A.; Chan, Y. L.; Ren, Z.; Wool, I. G.; Steitz, T. A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 13436–13441.
- (30) Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2006**, *34*, 697–708.
- (31) Eichman, B. F.; Vargason, J. M.; Mooers, B. H. M.; Ho, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3971–3976.
- (32) Shishkin, O. V.; Palamarchuk, G. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 4413–4422.
- (33) Palamarchuk, G. V.; Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2009**, *26*, 653–662.
- (34) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 0618–0622.
- (35) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- (36) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.
- (37) Cizek, J. *J. Chem. Phys.* **1966**, *45*, 4256.
- (38) Paldus, J.; Shavitt, I.; Cizek, J. *Phys. Rev. A* **1972**, *5*, 50–67.
- (39) Pople, J. A.; Headgordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968–5975.
- (40) Klein, D. J.; Moore, P. B.; Steitz, T. A. *J. Mol. Biol.* **2004**, *340*, 141–177.
- (41) Banas, P.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, *114*, 8701–8712.
- (42) Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, *114*, 6642–6652.
- (43) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (44) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (45) Franci, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (46) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.
- (47) Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.
- (48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision D.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (49) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (50) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340.
- (51) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555–569.
- (52) Burda, J. V.; Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670–9677.
- (53) Lipparini, F.; Scalmani, G.; Mennucci, B. *Phys. Chem. Chem. Phys.* **2009**, *11*, 11617–11623.
- (54) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (55) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (56) Boys, S. F.; Bernardi, F. *Mol. Phys.* **2002**, *100*, 65–73.
- (57) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.
- (58) Werner, H. J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *Molpro Version 2006.1*, a package of ab initio programs, 2006.
- (59) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319–325.
- (60) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464–470.
- (61) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778–784.
- (62) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 14103.
- (63) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103.
- (64) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (65) Adamo, C. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (66) van Leeuwen, R.; Baerends, E. J. *Phys. Rev. A* **1994**, *49*, 2421–2431.
- (67) Gruning, M.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **2001**, *114*, 652–660.
- (68) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (69) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (70) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comput. Chem.* **1983**, *4*, 294–301.
- (71) Gill, P. M. W.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1992**, *197*, 499–505.
- (72) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265–3269.
- (73) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- (74) Eichkorn, K.; Treutler, O.; Ohm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652–660.
- (75) Case, D. A.; Darden, T. A.; Chetham, III, T. E.; Simmerling, C. E.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *Amber 9*; University of California: San Francisco, 2006.
- (76) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (77) Banas, P.; Otyepka, M.; Jerabek, P.; Petrek, M.; Damborsky, J. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 375–383.
- (78) Ditzler, M. A.; Sponer, J.; Walter, N. G. *RNA* **2009**, *15*, 560–575.
- (79) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.
- (80) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.; Yajima, R. *Biopolymers* **2004**, *73*, 90–109.
- (81) Sponer, J.; Berger, I.; Spackova, N.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **2000**, 383–407.
- (82) Fadrna, E.; Spackova, N.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E.; Kulinski, T.; Sponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 2514–2530.
- (83) Zgarbova, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476–10493.
- (84) Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 1524–1544.
- (85) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136–1150.

- (86) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. *J. Comput. Chem.* **2009**, *30*, 2187–2193.
- (87) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. *Acc. Chem. Res.* **2010**, *43*, 40–47.
- (88) Kuzmin, Y. I.; Da Costa, C. P.; Cottrell, J. W.; Fedor, M. J. *J. Mol. Biol.* **2005**, *349*, 989–1010.
- (89) Cottrell, J. W.; Kuzmin, Y. I.; Fedor, M. J. *J. Biol. Chem.* **2007**, *282*, 13498–13507.
- (90) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferre-D'Amare, A. R. *Science* **2002**, *298*, 1421–1424.
- (91) Torelli, A. T.; Krucinska, J.; Wedekind, J. E. *RNA* **2007**, *13*, 1052–1070.
- (92) MacElrevey, C.; Salter, J. D.; Krucinska, J.; Wedekind, J. E. *RNA* **2008**, *14*, 1600–1616.
- (93) Guo, M.; Spitale, R. C.; Volpini, R.; Krucinska, J.; Cristalli, G.; Carey, P. R.; Wedekind, J. E. *J. Am. Chem. Soc.* **2009**, *131*, 12908–12909.
- (94) Sponer, J.; Spackova, N. *Methods* **2007**, *43*, 278–290.
- (95) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669–8681.
- (96) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Sponer, J. *Biophys. J.* **2006**, *91*, 626–638.
- (97) Barrick, J. E.; Corbino, K. A.; Winkler, W. C.; Nahvi, A.; Mandal, M.; Collins, J.; Lee, M.; Roth, A.; Sudarsan, N.; Jona, I.; Wickiser, J. K.; Breaker, R. R. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6421–6426.
- (98) Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. *Nature* **2004**, *428*, 281–286.
- (99) Milewski, S. *BBA-Protein Struct. M.* **2002**, *1597*, 173–192.
- (100) Collins, J. A.; Irnov, I.; Baker, S.; Winkler, W. C. *Genes Dev.* **2007**, *21*, 3356–3368.
- (101) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.
- (102) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023–5063.